

机器学习导论

习题六

151250104, 卢以宁, kiwiloveskiwis@gmail.com

2017 年 6 月 7 日

1 [20pts] Ensemble Methods

- (1) [10pts] 试说明 Boosting 的核心思想是什么, Boosting 中什么操作使得基分类器具备多样性?
- (2) [10pts] 试析随机森林为何比决策树 Bagging 集成的训练速度更快。

Solution. (1) Boosting 的核心思想是串行对基学习期进行训练, 在训练某个基学习期时, 根据在此之前的基学习器的表现调整样本分布 (即, 预测错误的样例获得更多关注), 从而让当前学习器的训练更有针对性。其中, 改变样本分布的做法使得基学习期具备多样性。

(2) 这是因为随机森林每次只选择一部分属性进行训练。

2 [20pts] Bagging

考虑一个回归学习任务 $f: \mathbb{R}^d \rightarrow \mathbb{R}$ 。假设我们已经学得 M 个学习器 $\hat{f}_1(\mathbf{x}), \hat{f}_2(\mathbf{x}), \dots, \hat{f}_M(\mathbf{x})$ 。我们可以将学习器的预测值看作真实值项加上误差项

$$\hat{f}_m(\mathbf{x}) = f(\mathbf{x}) + \epsilon_m(\mathbf{x}) \quad (2.1)$$

每个学习器的期望平方误差为 $\mathbb{E}_{\mathbf{x}}[\epsilon_m(\mathbf{x})^2]$ 。所有的学习器的期望平方误差的平均值为

$$E_{av} = \frac{1}{M} \sum_{m=1}^M \mathbb{E}_{\mathbf{x}}[\epsilon_m(\mathbf{x})^2] \quad (2.2)$$

M 个学习器得到的 Bagging 模型为

$$\hat{f}_{bag}(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^M \hat{f}_m(\mathbf{x}) \quad (2.3)$$

Bagging 模型的误差为

$$\epsilon_{bag}(\mathbf{x}) = \hat{f}_{bag}(\mathbf{x}) - f(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^M \epsilon_m(\mathbf{x}) \quad (2.4)$$

其期望平均误差为

$$E_{bag} = \mathbb{E}_{\mathbf{x}}[\epsilon_{bag}(\mathbf{x})^2] \quad (2.5)$$

- (1) [10pts] 假设 $\forall m \neq l, \mathbb{E}_{\mathbf{x}}[\epsilon_m(\mathbf{x})] = 0, \mathbb{E}_{\mathbf{x}}[\epsilon_m(\mathbf{x})\epsilon_l(\mathbf{x})] = 0$ 。证明

$$E_{bag} = \frac{1}{M} E_{av} \quad (2.6)$$

- (2) [10pts] 试证明不需对 $\epsilon_m(\mathbf{x})$ 做任何假设, $E_{bag} \leq E_{av}$ 始终成立。(提示: 使用 Jensen's inequality)

Proof. (1)

$$\begin{aligned} E_{bag} &= \mathbb{E}_{\mathbf{x}}[\epsilon_{bag}(\mathbf{x})^2] \\ &= \frac{1}{M^2} \mathbb{E}_{\mathbf{x}}[(\sum_{m=1}^M \epsilon_m(\mathbf{x}))^2] \\ &= \frac{1}{M^2} \sum_{m=1}^M \mathbb{E}_{\mathbf{x}}[(\epsilon_m(\mathbf{x}))^2] \\ &= \frac{1}{M} E_{av} \end{aligned} \quad (2.7)$$

□

- (2) 因为 $f(x) = x^2$ 是凸函数, 故由 Jensen's inequality, $f(\frac{1}{n}(\sum_{i=1}^n x_i)) \leq \frac{1}{n} \sum_{i=1}^n f(x_i)$ 则

$$\begin{aligned} E_{bag} &= \mathbb{E}_{\mathbf{x}}[\frac{1}{M}(\sum_{m=1}^M \epsilon_m(\mathbf{x}))^2] \\ &\leq \frac{1}{M} \sum_{m=1}^M \mathbb{E}_{\mathbf{x}}[(\epsilon_m(\mathbf{x}))^2] \\ &= E_{av} \end{aligned} \quad (2.8)$$

3 [30pts] AdaBoost in Practice

- (1) [25pts] 请实现以 Logistic Regression 为基分类器的 AdaBoost, 观察不同数量的 ensemble 带来的影响。详细编程题指南请参见链接: http://lamda.nju.edu.cn/ml2017/PS6/ML6_programming.html
- (2) [5pts] 在完成上述实践任务之后, 你对 AdaBoost 算法有什么新的认识吗? 请简要谈谈。

Solution. (1) 可能是因为 LR 并不适合做 adaboost 的基分类器

- (2) 线性分类器的线性加权还是线性的, 没有提高分类器的 VC 维 (也就是分类性能的量级)
- (3) 决策树桩可能比较合适
- (4) 那岂不是爆炸了¹

¹"Boosting shines when there is no terse functional form around": <https://stats.stackexchange.com/questions/186966/gradient-boosting-for-linear-regression-why-does-it-not-work>